



# Minería de datos y Patrones

Version 2024-I

## LDA, QDA, Mahalanobis

[ Capítulo 4 ]

**Dr. José Ramón Iglesias**

DSP-ASIC BUILDER GROUP

Director Semillero TRIAC

Ingeniería Electronica

Universidad Popular del Cesar

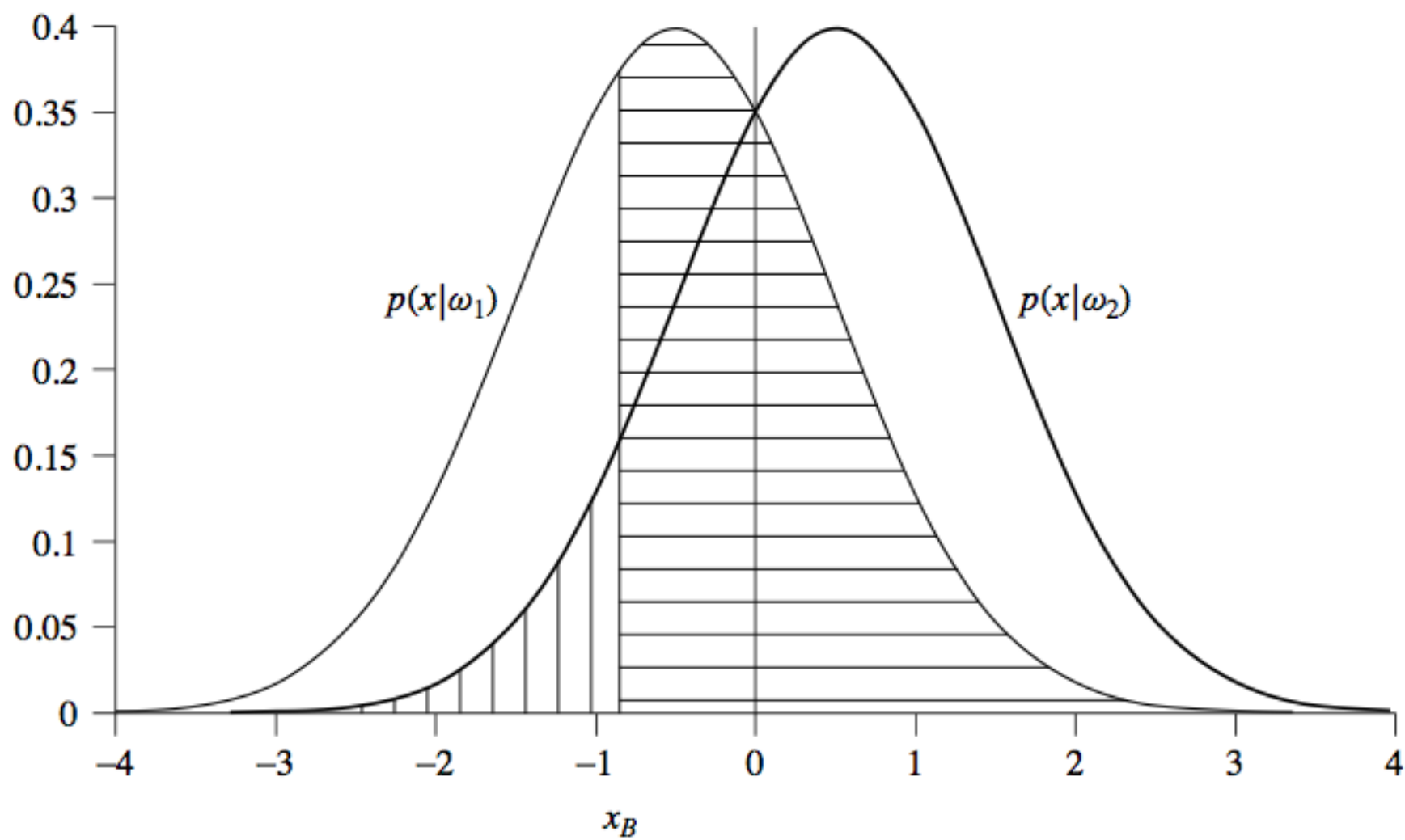
Clasificador de Bayes:  $\mathbf{x}$  es clasificado como clase  $j$  si

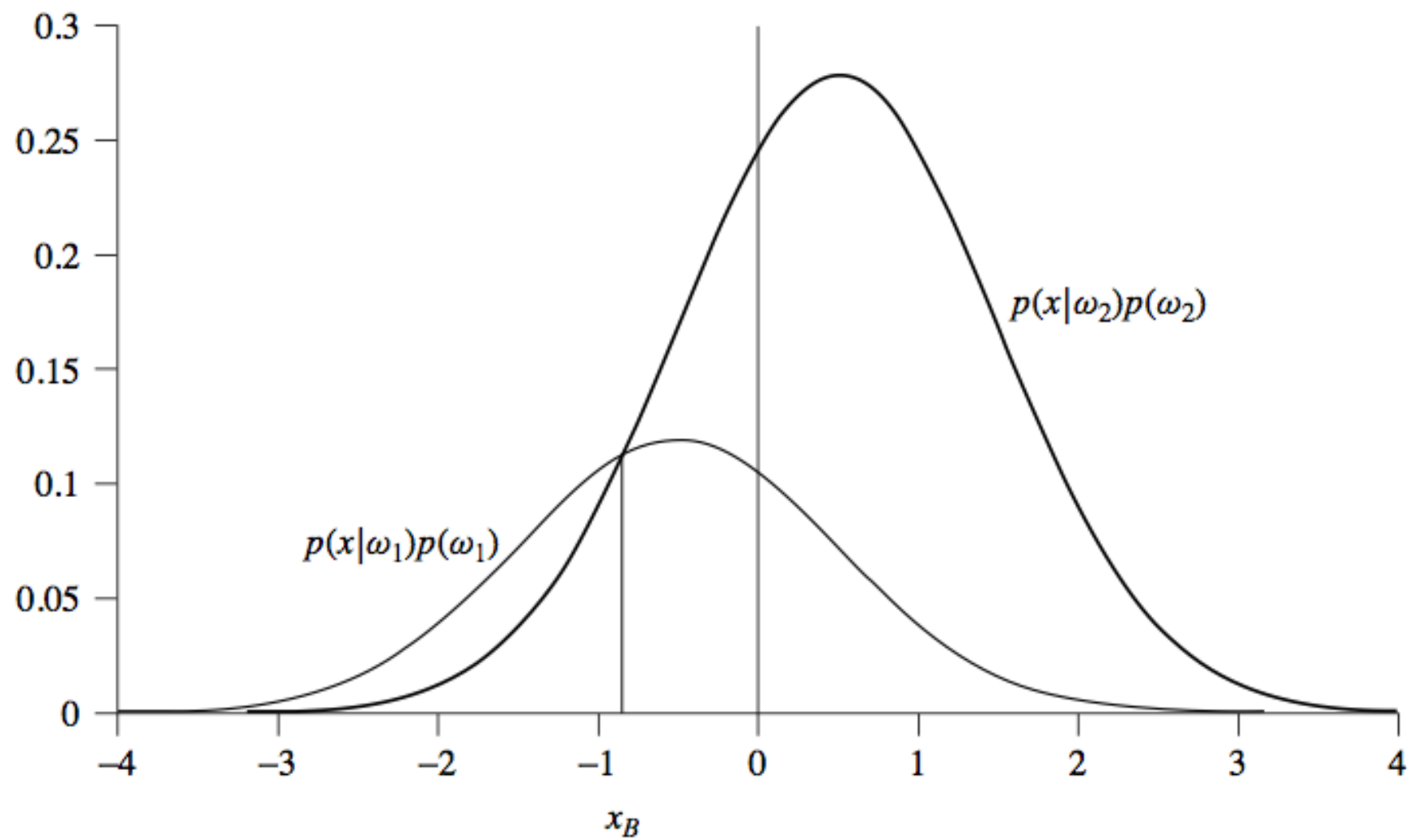
$$p(\omega_j|\mathbf{x}) > p(\omega_k|\mathbf{x}) \quad k = 1, \dots, C; k \neq j$$

Usando el teorema de Bayes:

$$p(\omega_i|\mathbf{x}) = p(\omega_i) \frac{p(\mathbf{x}|\omega_i)}{p(\mathbf{x})}$$

$$p(\mathbf{x}|\omega_j)p(\omega_j) > p(\mathbf{x}|\omega_k)p(\omega_k) \quad k = 1, \dots, C; k \neq j$$





Para distribuciones Gaussianas

$$p(\mathbf{x}|\omega_i) = \frac{1}{(2\pi)^{p/2}|\Sigma_i|^{1/2}} \exp \left\{ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \Sigma_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i) \right\}$$

Estimador de Matriz de Covarianza:  $\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^T$

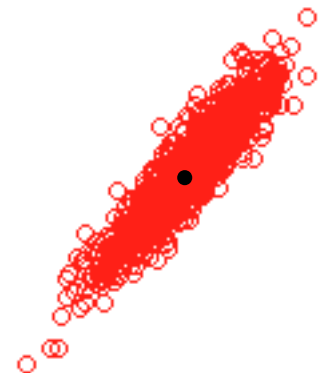
Ejemplos de  $\Sigma$  y  $\mu$  en 2D:



6.0057	-0.1020
-0.1020	1.0632



7.7947	-0.7772
-0.7772	43.1459



15.1951	21.8267
21.8267	37.6734

## Clasificador de Bayes

$$p(\omega_j|\mathbf{x}) > p(\omega_k|\mathbf{x}) \quad k = 1, \dots, C; k \neq j$$

$$p(\mathbf{x}|\omega_j)p(\omega_j) > p(\mathbf{x}|\omega_k)p(\omega_k) \quad k = 1, \dots, C; k \neq j$$

$$\log\{p(\mathbf{x}|\omega_j)p(\omega_j)\} > \log\{p(\mathbf{x}|\omega_k)p(\omega_k)\} \quad k = 1, \dots, C; k \neq j$$

Para distribuciones Gaussianas:

$$p(\mathbf{x}|\omega_i) = \frac{1}{(2\pi)^{p/2} |\mathbf{\Sigma}_i|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_i)^T \mathbf{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) \right\}$$

$$\begin{aligned} \log(p(\mathbf{x}|\omega_i) p(\omega_i)) &= \log(p(\mathbf{x}|\omega_i)) + \log(p(\omega_i)) \\ &= -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_i)^T \mathbf{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) - \frac{1}{2} \log(|\mathbf{\Sigma}_i|) \\ &\quad - \frac{p}{2} \log(2\pi) + \log(p(\omega_i)) \end{aligned}$$

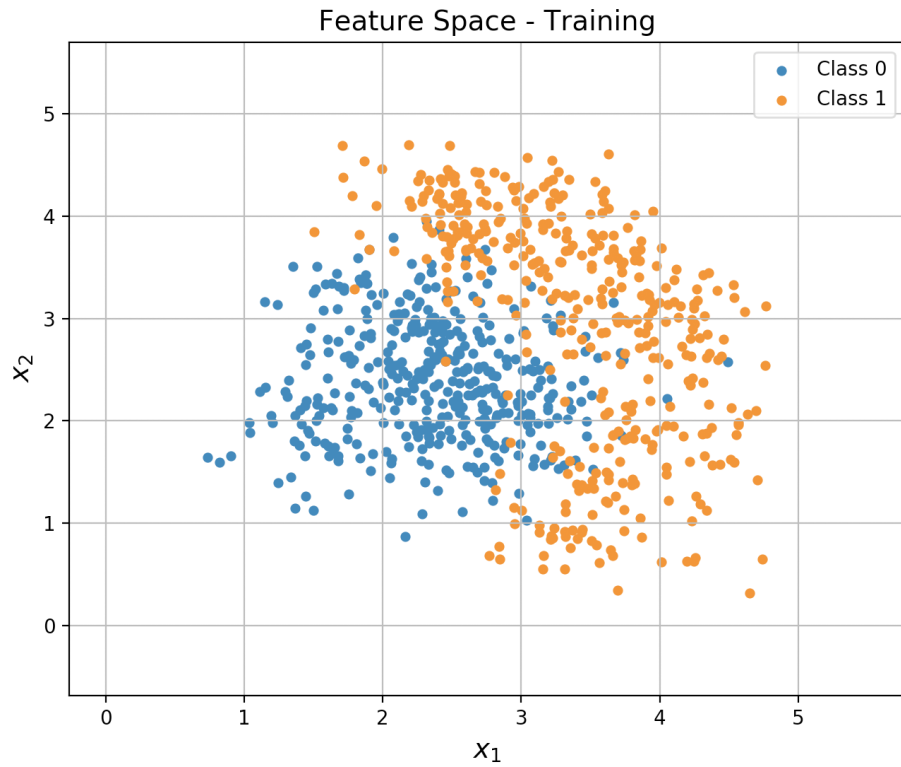
$$\log\{p(\mathbf{x}|\omega_j)p(\omega_j)\} > \log\{p(\mathbf{x}|\omega_k)p(\omega_k)\} \quad k = 1, \dots, C; k \neq j$$

$$\begin{aligned} \log(p(\mathbf{x}|\omega_i) p(\omega_i)) &= \log(p(\mathbf{x}|\omega_i)) + \log(p(\omega_i)) \\ &= -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) - \frac{1}{2} \log(|\boldsymbol{\Sigma}_i|) \\ &\quad - \cancel{\frac{p}{2} \log(2\pi)} + \log(p(\omega_i)) \end{aligned}$$

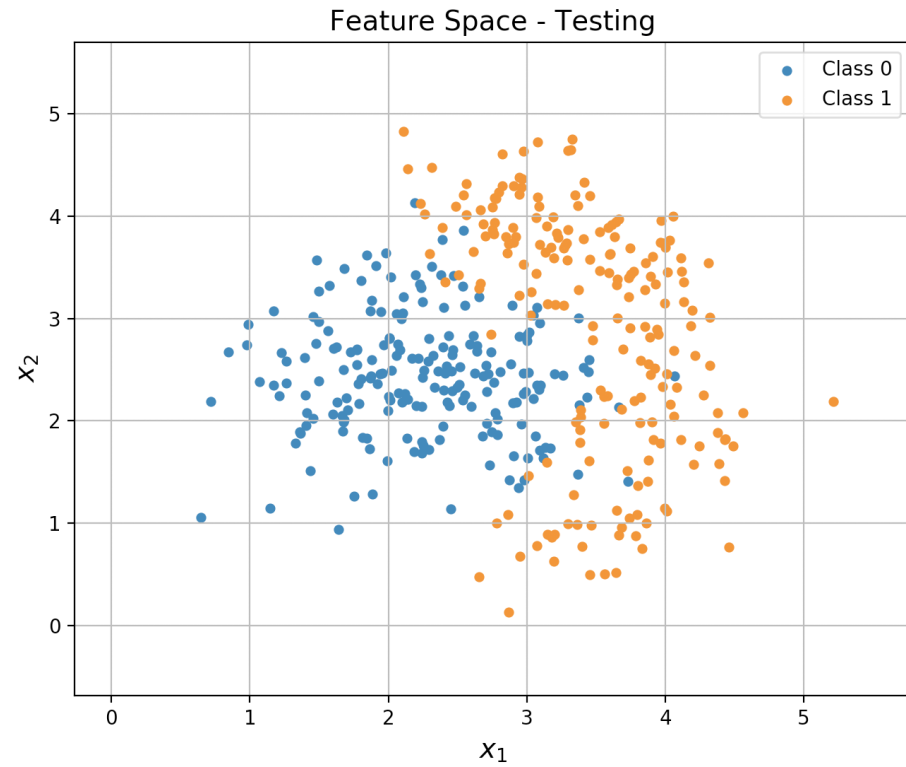
Este término es constante, se puede eliminar de la desigualdad.

# Ejemplo

$(X, d)$   
TRAINING



$(X_t, d_t)$   
TESTING





# LDA

$$\log\{p(\mathbf{x}|\omega_j)p(\omega_j)\} > \log\{p(\mathbf{x}|\omega_k)p(\omega_k)\} \quad k = 1, \dots, C; k \neq j$$

$$\begin{aligned}\log(p(\mathbf{x}|\omega_i) p(\omega_i)) &= \log(p(\mathbf{x}|\omega_i)) + \log(p(\omega_i)) \\ &= -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) - \frac{1}{2} \log(|\boldsymbol{\Sigma}_i|) \\ &\quad - \frac{p}{2} \log(2\pi) + \log(p(\omega_i))\end{aligned}$$

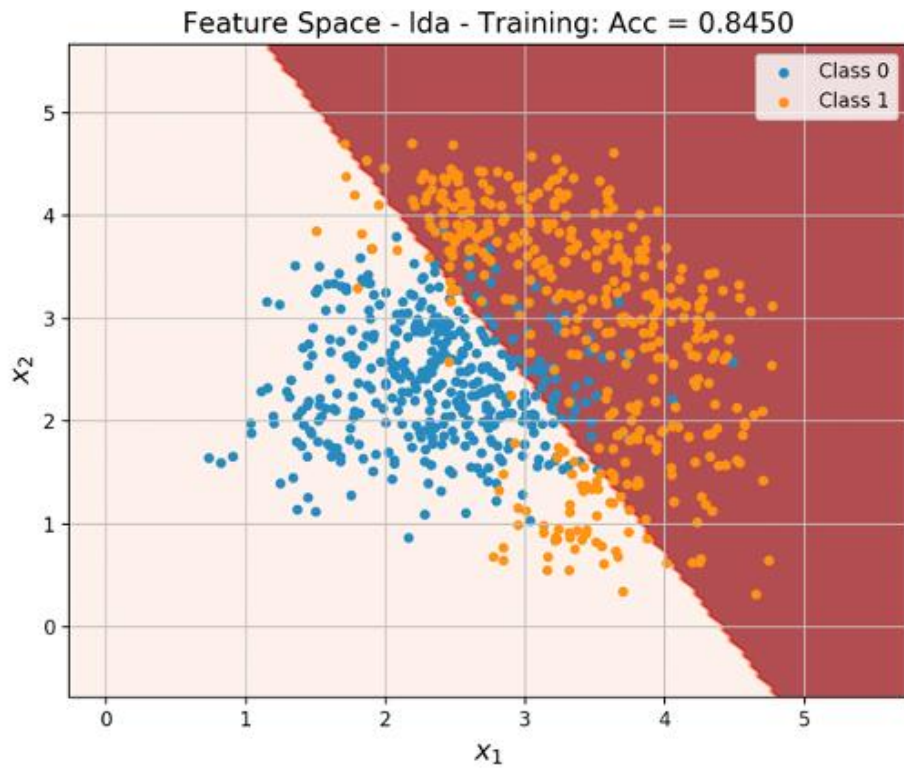
En LDA (Análisis Discriminante Lineal) se supone que  $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$  (es constante) .

$\boldsymbol{\Sigma}$  se calcula a partir de datos de entrenamiento. Una buena estimación es el promedio de las matrices de covarianza individuales:  $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2) / 2$

# Ejemplo

## LDA

$(X, d)$   
TRAINING



$(X_t, d_t)$   
TESTING



# MAHALANOBIS

$$\log\{p(\mathbf{x}|\omega_j)p(\omega_j)\} > \log\{p(\mathbf{x}|\omega_k)p(\omega_k)\} \quad k = 1, \dots, C; k \neq j$$

$$\begin{aligned} \log(p(\mathbf{x}|\omega_i) p(\omega_i)) &= \log(p(\mathbf{x}|\omega_i)) + \log(p(\omega_i)) \\ &= -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) - \frac{1}{2} \log(|\boldsymbol{\Sigma}_i|) \\ &\quad - \frac{p}{2} \log(2\pi) + \log(p(\omega_i)) \end{aligned}$$

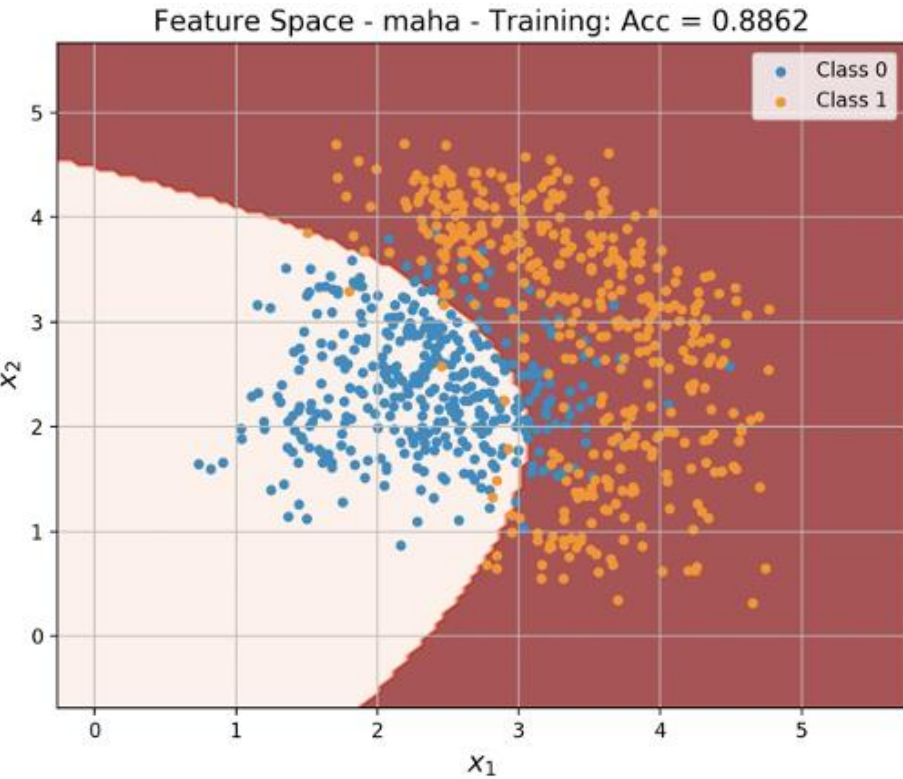
En el Clasificador Mahalanobis se asume  $p(\omega_i) = p$  y

las matrices  $\boldsymbol{\Sigma}_i$  son distintas.

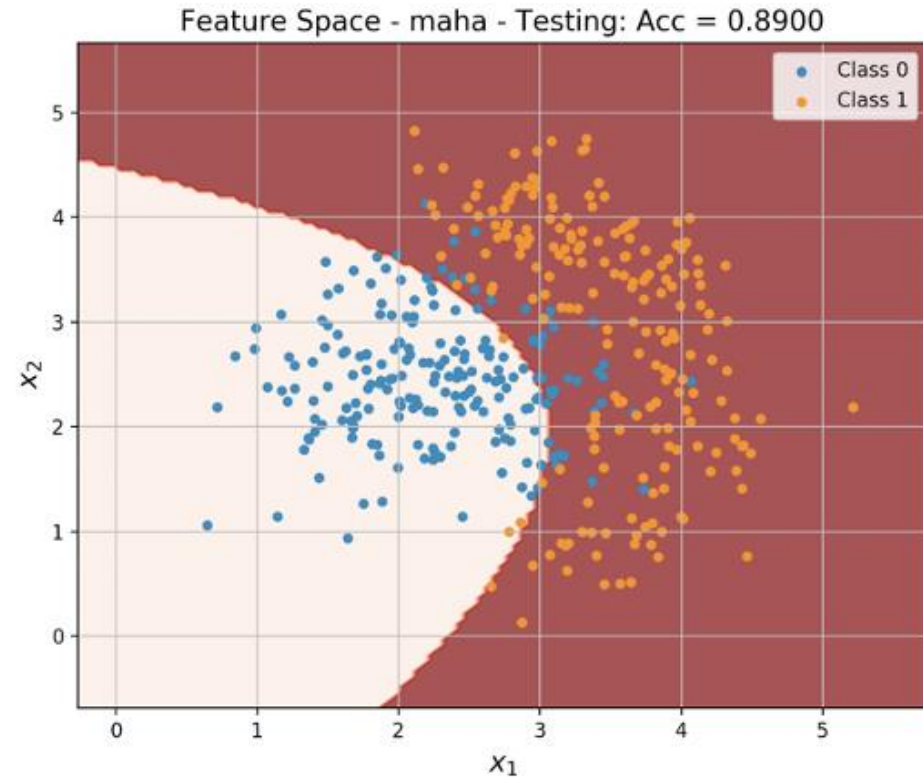
# Ejemplo

# Mahalanobis

$(X, d)$   
TRAINING



$(X_t, d_t)$   
TESTING



# MAHALANOBIS-0

$$\log\{p(\mathbf{x}|\omega_j)p(\omega_j)\} > \log\{p(\mathbf{x}|\omega_k)p(\omega_k)\} \quad k = 1, \dots, C; k \neq j$$

$$\begin{aligned} \log(p(\mathbf{x}|\omega_i) p(\omega_i)) &= \log(p(\mathbf{x}|\omega_i)) + \log(p(\omega_i)) \\ &= -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) - \frac{1}{2} \log(|\boldsymbol{\Sigma}_i|) \\ &\quad - \frac{p}{2} \log(2\pi) + \log(p(\omega_i)) \end{aligned}$$

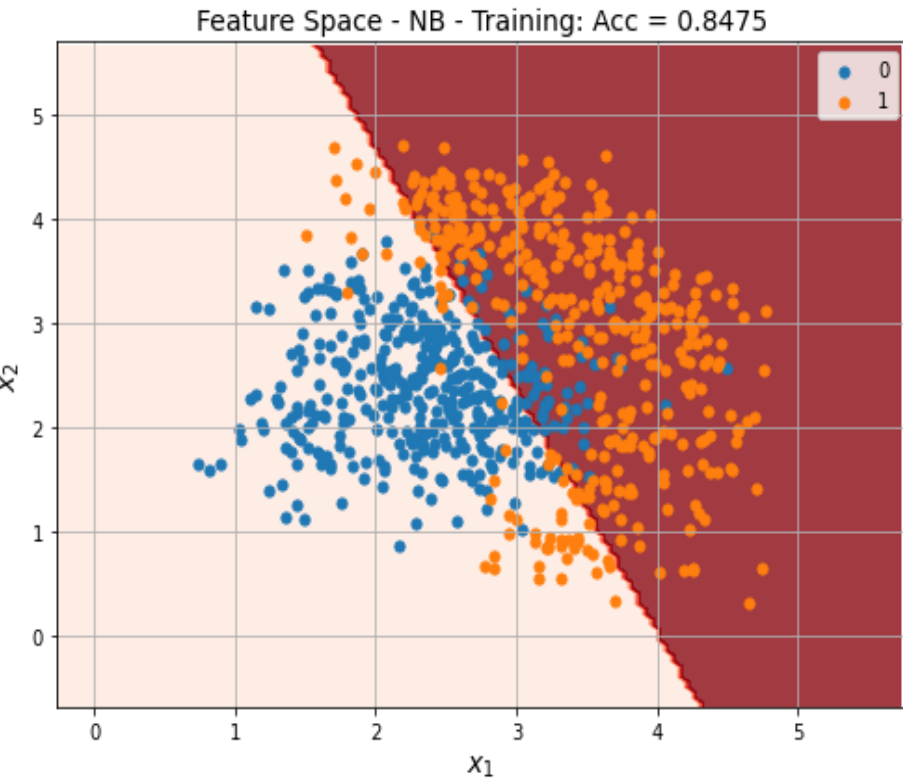
En el Clasificador Mahalanobis se asume  $p(w_i) = p$ .

Hay una variante de Mahalanobis en la que se supone que  $\Sigma_i = \Sigma$ .

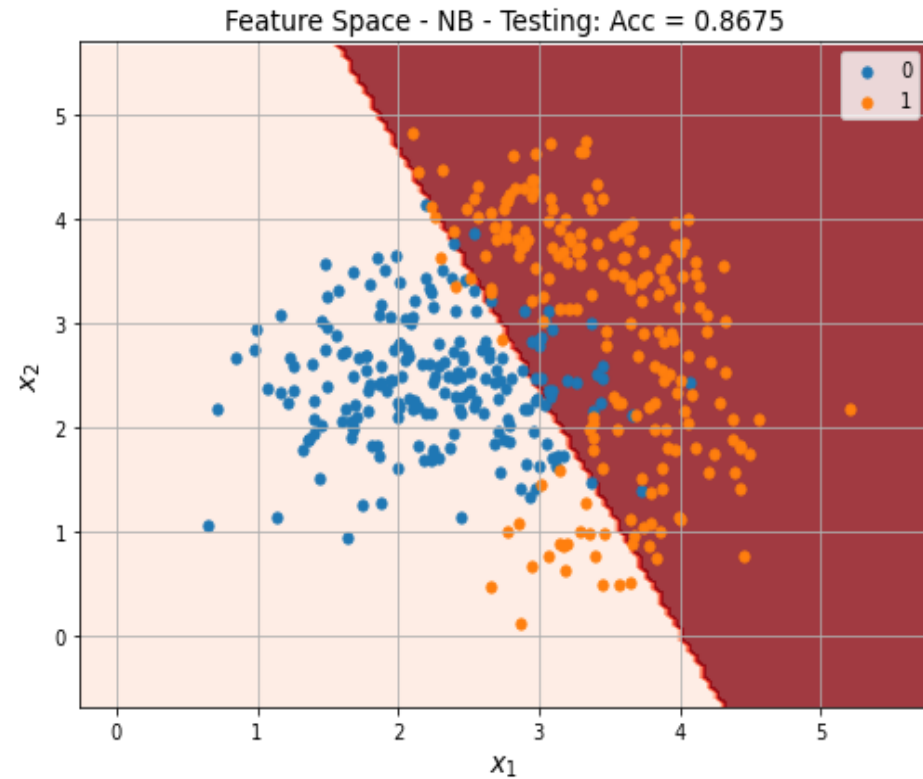
# Ejemplo

Mahalanobis-0 ( $\Sigma_i = \Sigma$ )

(X,d)  
TRAINING



(X<sub>t</sub>,d<sub>t</sub>)  
TESTING



# QDA

$$\log\{p(\mathbf{x}|\omega_j)p(\omega_j)\} > \log\{p(\mathbf{x}|\omega_k)p(\omega_k)\} \quad k = 1, \dots, C; k \neq j$$

$$\begin{aligned}\log(p(\mathbf{x}|\omega_i)p(\omega_i)) &= \log(p(\mathbf{x}|\omega_i)) + \log(p(\omega_i)) \\ &= -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i) - \frac{1}{2} \log(|\boldsymbol{\Sigma}_i|) \\ &\quad - \frac{p}{2} \log(2\pi) + \log(p(\omega_i))\end{aligned}$$

En QDA (Análisis Discriminante Cuadrático) se supone que  $\boldsymbol{\Sigma}_i$  y  $p(\omega_i)$  son diferentes.

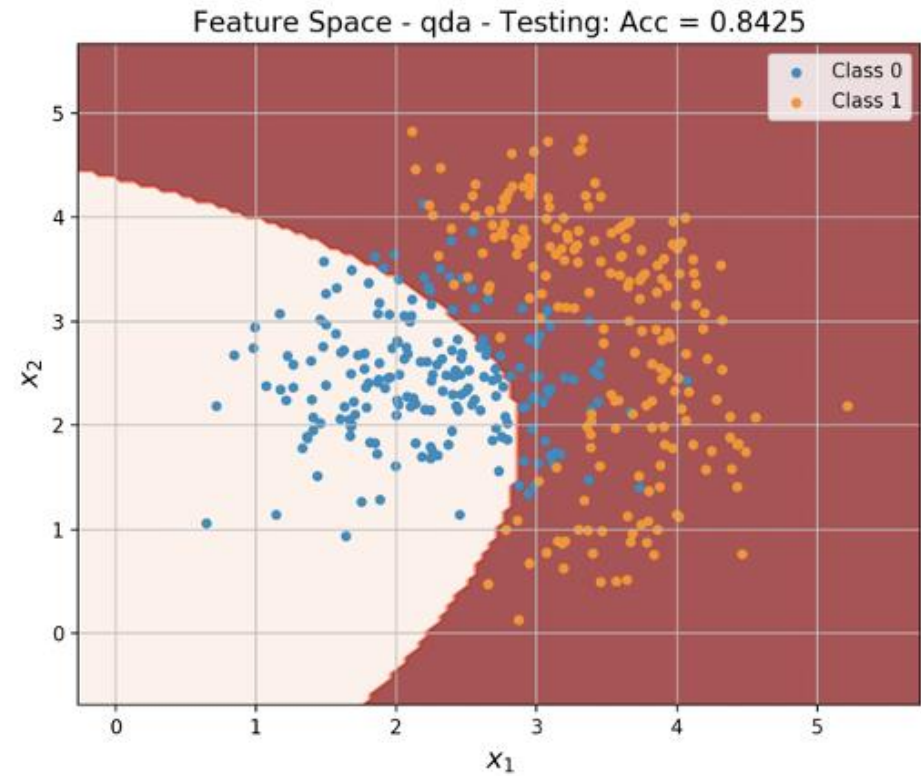
# Ejemplo

## QDA

$(X, d)$   
TRAINING



$(X_t, d_t)$   
TESTING





$$\log(p(\mathbf{x}|\omega_i)) = -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i) - \frac{1}{2} \log(|\boldsymbol{\Sigma}_i|) - \frac{p}{2} \log(2\pi) + \log(p(\omega_i))$$

LDA

$$= -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i) - \cancel{\frac{1}{2} \log(|\boldsymbol{\Sigma}_i|)} - \cancel{\frac{p}{2} \log(2\pi)} + \log(p(\omega_i))$$

$\Sigma_i = \Sigma = \text{cte}$        $\text{cte}$

Mahalanobis

$$= -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i) - \cancel{\frac{1}{2} \log(|\boldsymbol{\Sigma}_i|)} - \cancel{\frac{p}{2} \log(2\pi)} + \cancel{\log(p(\omega_i))}$$

$\Sigma_i = \Sigma = \text{cte}$        $\text{cte}$        $p(\omega_i) = \text{cte}$

QDA

$$= -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i) - \frac{1}{2} \log(|\boldsymbol{\Sigma}_i|) - \cancel{\frac{p}{2} \log(2\pi)} + \log(p(\omega_i))$$

$\text{cte}$